STATISTICAL INFERENCE ABOUT MARKOV CHAINS

by

SAI-SING LIN

B. Sc., Taiwan Normal University, 1958

A MASTER'S REPORT

submitted in partial fulfillment of the

requirement for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

Approved by:

A. M. Feyerherm

Major Professor

TABLE OF CONTENTS

INTRODUCTION

The subject of this report is a relatively new topic in probability and statistics. Most of the papers dealing with statistical inference in Markov chains appeared during the past fifteen years. Markov chains are defined in standard texts on probability and stochastic processes such as Feller (1957) and Parzen (1961).

Anderson (1957) studied statistical inference in Markov chains for the case of repeated observations on the same chain. In this circumstance, each observation is a sequence of states, over a finite number T of time points, from a Markov chain with common transition probability matrices $P_t = \left\{p_{ij}(t)\right\}$, and $n_i(0)$ observations are in state i at time zero. Under the assumption that $n_i(0) \to \infty$, he presented likelihood ratio tests for the following hypotheses: (a) P_t is stationary (i.e., $P_t = P = \{p_{ij}\}$) against the alternative that it varies over time, (b) P is a given matrix against the alternative that it is not, (c) the process is 1st-order against the alternative that it is 2nd-order.

Anderson and Goodman (1957) presented χ^2-tests for goodness of fit which are analogous to the likelihood ratio test for the three hypothesis just stated and for the following: (d) the process is rth-order against the alternative that it is (r+1)st-order. (f) the s samples of observations are samples from the same Markov chain P.

Problems of estimation of transition probabilities, testing of goodness of fit and order of a chain were studied by Bartlett (1951) and Hoel (1954) for the situation where only a single sequence of states is observed. They considered the asymptotic theory as the number of time points increase.

The work of Anderson and Goodman (1957) has promoted application of the

theory of Markov chains in a number of different disciplines and it is their work which will be reported in some detail.

An example by Anderson (1954) in which he applied statistical methods to the problem of studying voter intentions introduced repeated Markov chains to social scientists. Gabriel and Neumann (1957) and Feyerherm and Bark (1965) have applied the theory to the study of precipitation patterns.

ESTIMATION OF THE PARAMETERS OF A 1ST-ORDER MARKOV CHAIN

Definition and Notation

Consider a sequence of observations in which each observation can be in any one of m distinct states at a discrete time point t. Let $p_{ij}(t)$ $(i,j = 1,2,\cdots,m;\ t = 1,2,\cdots,T)$ be the probability of state j at time t, given state i at time (t-1). The transition probability matrices are defined by

$$P_t = \left\{ p_{ij}(t) \right\} \tag{1.1}$$

where

(1) $p\ \ (t) \geqq 0$; for all (i,j) and t,

(2) $\sum_{j=1}^{m} p_{ij}(t) = 1$; for all i, and t,

(3) $p_{ij}(t) = \sum_{k} p_{ik}(t') \ p_{kj}(t)$; for any times t $t' \geq 0$ and states i and j

The probability law of a homogeneous Markov chain P is completely determined once one knows the transition probability matrices given by $P_t = \left\{ p_{ij}(t) \right\}$ and the unconditional probability vector $p(0) = \left\{ p_i(0) \right\}$ at time zero (see Parzen, 1962, p.196).

Model

Assume that there are $n_i(0)$ individuals in state i at t=0. An observation on a given individual consists of the sequence of states that the individual is in at $t = 0,1,\cdots,T$ namely i(0), i(1), ..., i(T). If the initial state i(0) is given, then there are m^T possible sequences. For a 1st-order Markov chain, these represent mutually exclusive events with probabilities

$$p_{i(0)i(1)\cdots i(T)} = p_{i(0)i(1)}\, p_{i(1)i(2)} \cdots p_{i(T-1)i(T)} \tag{2.1}$$

when the transition probabilities are stationary.

If they are not stationary, then

$$p_{i(0)i(1)\cdots i(T)}{}^{(T)} = p_{i(0)i(1)}{}^{(1)}\, p_{i(1)i(2)}{}^{(2)} \cdots p_{i(T-1)i(T)}{}^{(T)}.$$

Let $n_{ij}(t)$ = no. of individuals in state i at time (t-1) and i at time t.

and $n_{i(0)i(1)\cdots i(T)}$ be the number of individuals whose sequence of states is i(0), i(1), ... i(t). Then

$$n_{gj}(t) = \Sigma\, n_{i(0)i(1)\cdots i(T)} \tag{2.2}$$

where the sum is over all values of the i's with i(t-1) = g and i(t) = j. The probability, in the nmT dimensional space describing all sequences for all n individuals (for each initial state there are nT dimensions) of a given ordered set of sequences for the n individuals is:

$$\prod_{i(0)\cdots i(T)} \left[p_{i(0)i(1)\cdots i(T)}{}^{(t)} \right]^{n_{i(0)i(1)\cdots i(T)}}$$

$$= \prod_{i(0)\cdots i(T)} \left[p_{i(0)i(1)}{}^{(1)}\, p_{i(1)i(2)}{}^{(2)} \cdots p_{i(T-1)i(T)}{}^{(T)} \right]^{n_{i(0)\cdots i(T)}}$$

$$= \left\{ \prod_{i(0)\cdots i(T)} \left[p_{i(0)i(1)}{}^{(1)} \right]^{n_{i(0)i(1)}{}^{(1)}} \right\} \cdots \left\{ \prod_{i(0)\cdots i(T)} \left[p_{i(T-1)i(T)}{}^{(T)} \right]^{n_{i(T-1)i(T)}{}^{(T)}} \right\}$$

$$=\left\{\prod_{i(0)i(1)}\left[p_{i(0)i(1)}(1)\right]^{n_{i(0)i(1)}(1)}\right\}\cdots$$

$$\left\{\prod_{i(T-1)i(T)} p_{i(T-1)i(T)}(T)^{n_{i(T-1)i(T)}(T)}\right\}$$

$$=\prod_{t=1}^{T}\prod_{g,j} p_{gj}(t)^{n_{gj}(t)}. \tag{2.3}$$

Therefore, according to "factor theorem" (Hogg and Craig, 1965) the set of numbers $n_{ij}(t)$ form a set of sufficient statistics.

Let $\quad n_i(t-1) = \sum_{j=1}^{m} n_{ij}(t).$

Then the conditional distribution of $n_{ij}(t)$, $j = 1,2,\cdots,m$, given $n_i(t-1)$ is

$$\frac{n_i(t-1)!}{\prod_{j-1}^{m} n_{ij}(t)!}\prod_{j=1}^{m} p_{ij}(t)^{n_{ij}(t)}.$$

The distribution of $n_{ij}(t)$ (conditional on $n_i(0)$) is

$$\prod_{t=1}^{T}\left\{\prod_{j=1}^{m}\left[\frac{n_i(t-1)!}{\prod_{j=1}^{m} n_{ij}(t)!}\prod_{j=1}^{m} p_{ij}(t)^{n_{ij}(t)}\right]\right\}.$$

If the transition probabilities are stationary, then the set $n_{ij} = \sum_{t=1}^{T} n_{ij}(t)$ can be seen to be a set of sufficient statistics and (2.3) can be written in the form

$$\prod_{t=1}^{T} \prod_{g,j} p_{gj}(t)^{n_{gj}(t)} = \prod_{i,j} p_{ij}^{n_{ij}}. \tag{2.4}$$

Maximum Likelihood Estimates

The stationary transition probabilities p_{ij} can be estimated by maximizing the probability (2.4) with respect to the p_{ij} under the conditions

(1) $p_{ij} \geq 0$; $i,j = 1,2,\cdots,m$,

(2) $\sum_{j=1}^{m} p_{ij} = 1$; for all i, where the n_{ij} are the actual observations.

For m independent samples, the i^{th} sample $(i = 1,2,\cdots,m)$ consists of $n_i^* = \Sigma_j n_{ij}$ multinomial trials with probabilities p_{ij} $(i,j = 1,2,\cdots,m)$. Then the maximum likelihood estimates for p_{ij} are

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i^*} = \sum_{t=1}^{T} n_{ij}(t) \Big/ \sum_{k=1}^{m} \sum_{t=1}^{T} n_{ik}(t)$$

$$= \sum_{t=1}^{T} n_{ij}(t) \Big/ \sum_{t=0}^{T-1} n_i(t).$$

When the transition probabilities are not necessarily stationary, the maximum likelihood estimates for the $p_{ij}(t)$ are

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t-1)}$$

$$= \frac{n_{ij}(t)}{\sum_{k=1}^{m} n_{ik}(t)} .$$

Formally the estimates are the same as one would obtain if for each i and t one had $n_i(t-1)$ observations on a multinomial distribution with probabilities $p_{ij}(t)$ and with resulting numbers $n_{ij}(t)$.

The estimates can be described in the following way: Let the entries $n_{ij}(t)$ for given t be entered in a two-way m x m table. The estimate of $p_{ij}(t)$ is the (i,j)th entry in the table divided by the sum of the entries in the i(th) row. To estimate p_{ij}, for a stationary chain, add the corresponding entries in the two-way tables for $t = 1,2,\cdots,T$ and obtain a two-way table with entries $n_{ij} = \sum_t n_{ij}(t)$. The estimate of p_{ij} is the (i,j)th entry of the table of n_{ij}'s divided by the sum of the entries in the i(th) row.

Asymptotic Behavior of $n_{ij}(t)$

Consider the following theorem:

Theorem: If $(x_{1\xi}, x_{2\xi}, \cdots, x_{k\xi}, \xi = 1,2,\cdots,n)$ is a sample of size n from the mutinomial distribution $M(1 ; p_1, p_2, \cdots, p_k)$, then the sample sums $(z_1, z_2, \cdots, z_k)$ have, as their asymtotic distribution for large n, the distribution $N(\{np_i\}, \| n(p_i \; {}_{ij} - p_i p_j)\|)$ where δ_{ij} is the Kronecker delta. (Wilks, 1963, p.259)

Proof: The p.f. of the multinomial distribution $M(1 ; p_1, p_2, \cdots, p_k)$ is

$$p(x_1, x_2, \cdots, x_k) = \frac{1!}{x_1!\, x_2! \cdots x_{k+1}!} p_1^{x_1} p_2^{x_2} \cdots p_{k+1}^{x_{k+1}}$$

where $x_{k+1} = 1 - x_1 - x_2 - \cdots - x_k$ and $p_{k+1} = 1 - p_1 - p_2 - \cdots - p_k$

Then the characteristic function of the multinomial distribution is

$$\varphi(t_1, \cdots, t_k) = \Sigma\, e^{it_1x_1 + it_2x_2 + \cdots + it_kx_k}\, p(x_1, x_2, \cdots, x_k)$$

$$= \Sigma \frac{1!}{x_1! \cdots x_k!} (p_1e^{it_1})^{x_1} \cdots (p_ke^{it_k})^{x_k} (p_{k+1})^{x_{k+1}}$$

$$= (p_1e^{it_1} + p_2e^{it_2} + \cdots + p_ke^{it_k} + p_{k+1}).$$

It follows that

$$\mu(x_i) = p_i$$

$$\sigma^2(x_i, x_j) = \sigma_{ij} = p_i\delta_{ij} - p_ip_j$$

Then, using the result (see wilks, 1963, P.258)

$$\lim_{n\to\infty} P\left(\frac{(z_i - n\mu_i)}{\sqrt{n}} \le y_i,\ i = 1,2,\cdots,k\right)$$

$$= \frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{k/2}} \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_k} \exp\left(-\frac{1}{2} \sum_{i,j=1}^{k} \sigma^{ij}\mu_i\mu_j\right) d\mu_1 \cdots d\mu_k,$$

it follows that

$$(z_1, \cdots, z_k) \sim N(\{np_i\}, \|n(p_i\delta_{ij} - p_ip_j)\|).$$

To find the asymptotic behavior of the $\hat{p}_{ij}$, first consider the $n_{ij}(t)$. For each i(0), the set $n_{i(0)i(1)\cdots i(T)}$ are simply multinomial variables with sample size $n_{i(0)}(0)$ and parameters $p_{i(0)i(1)}\ p_{i(1)i(2)} \cdots p_{i(T-1)i(T)}$, and hence are asymptotically normally distributed as the samples size increase. The $n_{ij}(t)$ are linear combinations of these multinomial variables, and are also asymptotically normally distributed.

Let $P = [p_{ij}]$ and $p_{ij}^{[t]}$ be the elements of the matrix P^t. Then $p_{ij}^{[t]}$ is the probability of state j at time t given state i at time 0. Let $n_{k;ij}(t)$ be the number of sequences including state k at time 0, i at time (t-1) and j at time t. Then

$$n_{ij}(t) = \sum_{k=1}^{m} n_{k;ij}(t).$$

The probability associated with $n_{k;ij}(t) = p_{ki}^{[t-1]}\ p_{ij}$ with a sample size of $n_k(0)$. Thus

$$E\left\{ n_{k;ij}(t) \right\} = n_k(0)\ p_{ki}^{[t-1]} \cdot p_{ij}$$

$$\operatorname{var}\left\{ n_{k;ij}(t) \right\} = n_k(0)\ p_{kj}^{[t-1]} \cdot p_{ij}\left[1 - p_{ki}^{[t-1]}\ p_{ij}\right]$$

$$\operatorname{cov}\left\{ n_{k;ij}(t), n_{k;gh}(t) \right\} = -n_k(0)\ p_{ki}^{[t-1]}\ p_{ij} p_{kg}^{[t-1]}\ p_{gh}$$

$$(i,j) \neq (g,h).$$

Consider $n_{k;ij}(t) - n_{k;i}(t-1)\ p_{ij}$, where $n_{k;i}(t-1) = \sum_j n_{k;ij}(t)$.

Then the conditional distribution of $n_{k;ij}$ (t) given $n_{k;i}$ (t-1) is multinomial with the probabilities p_{ij}. Thus,

$$E\left\{ n_{k;ij}(t) \mid n_{k;i}(t-1) \right\} = n_{k;i}(t-1)\, p_{ij},$$

$$E\left\{ n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij} \right\}$$

$$= E \cdot E\left\{ \left[n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij} \right] \mid n_{k;i}(t-1) \right\}$$

$= 0$, (see wilks, 1963, P.84) .

$$\mathrm{var}\left[n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij} \right]$$

$$= E\left[n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij} - 0 \right]^2$$

$$= E \cdot E\left\{ \left[n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij} \right]^2 \mid n_{k;i}(t-1) \right\}$$

$$= E \cdot n_{k;i}(t-1)\, p_{ij}(1 - p_{ij})$$

$$= n_k(0)\, p_{ki}^{[t-1]}\, p_{ij}\, (1 - p_{ij}),$$

$$\mathrm{cov}\left(n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij} \right)$$

$$= E\left[n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij} \right]\left[n_{k;ih}(t-1) - n_{k;i}(t-1)\, p_{ih} \right]$$

$$= E\cdot E\Big\{\big[n_{k;ij}\ (t-1) - n_{k;i}\ (t-1)\ p_{ij}\big]\big[n_{k;ih}\ (t-1) - n_{k;i}\ (t-1)\ p_{ih}\big]$$

$$\Big|\, n_{k;i}\ (t-1)\Big\}$$

$$= E\big[-n_{k;i}\ (t-1)\ p_{ij}p_{ih}\big]$$

$$= -n_k(0)\ p_{ki}{}^{[t-1]}p_{ij}p_{ih}\ ;\ j \neq h,$$

$$E\big[n_{k;ij}\ (t) - n_{k;i}\ (t-1)\ p_{ij}\big]\big[n_{k;gh}\ (t) - n_{k;g}\ (t-1)\ p_{gh}\big]$$

$$= EE\Big\{\big[n_{k;ij}\ (t) - n_{k;i}\ (t-1)\ p_{ij}\big]\big[n_{k;gh}\ (t) - n_{k;g}\ (t-1)\ p_{gh}\big]$$

$$\Big|\, n_{k;i}\ (t-1),\ n_{k;g}\ (t-1)\Big\}$$

$$= 0,$$

$$E\big[n_{k;ij}\ (t) - n_{k;i}\ (t-1)\ p_{ij}\big]\big[n_{k;gh}\ (t+r) - n_{k;g}\ (t+r-1)\ p_{gh}\big]$$

$$= EE\Big\{\big[n_{k;ij}\ (t) - n_{k;i}\ (t-1)\ p_{ij}\big]\big[n_{k;gh}\ (t+r) - n_{k;g}\ (t+r-1)\ p_{gh}\big]$$

$$\Big|\, n_{k;ij}\ (t),\ n_{k;g}\ (t+r-1),\ n_{k;i}\ (t-1)\Big\}$$

$$= 0. \qquad r > 0$$

Thus, the random variables $n_{k;ij}(t) - n_{k;ij}(t-1)\ p_{ij}$ for $j = 1,2,\cdots,m$ have zero mean and variance and covariances of multinomial variables with probabilities p and sample size $n_k(0)\ p_{ki}{}^{[t-1]}$. The variables $n_{k;ij}(t) - n_{k;i}(t-1)\ p_{ij}$ and $n_{k;gh}(s) - n_{k;g}(s-1)\ p_{gh}$ are uncorrelated if $t \neq s,\ i \neq g$.

Asymptotic Distribution of the Estimates

Consider

$$\sqrt{n}\ (p_{ij} - p_{ij}) = \sqrt{n}\left[\frac{\sum\limits_t n_{ij}(t)}{\sum\limits_{t=1}^{T} n_i(t-1)} - p_{ij}\right]$$

$$= \sqrt{n}\left[\frac{\sum\limits_{t=1}^{T}\left[n_{ij}(t) - p_{ij}n_i(t-1)\right]}{\sum n_i(t-1)}\right],$$

$$= \sqrt{n}\left[\frac{\sum\limits_{k=1}^{m}\sum\limits_{t=1}^{T}\left[n_{k;ij}(t) - p_{ij}\ n_{k;i}(t-1)\right]}{\sum\limits_{t=1}^{T} n_i(t-1)}\right].$$

Since $n_{k;ij}(t)$ is a multinomial variable with probabilities p_{ij}, then $n_{k;ij}(t)/n$ Converges in probability to its expected value, when $n_k(0)/n\ \ n_k$. Thus

$$\underset{n\to\infty}{p\lim}\ \frac{1}{n}\, n_{k;ij}\ (t) = \underset{n\to\infty}{p\lim}\ \frac{1}{n}\sum_{t=1}^{T} n_i\ (t-1)$$

$$= \lim_{n\to\infty} \frac{1}{n}\, E \sum_{t=1}^{T} n_i\ (t-1)$$

$$= \lim_{n\to\infty} \frac{1}{n}\, E \sum_{t=1}^{T} \sum_{k=1}^{m} n_{k;i}\ (t-1)$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{T} \sum_{k=1}^{m} E\left(n_{k;i}\ (t-1)\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{T} \sum_{k=1}^{m} n_k\ (0)\ p_{ki}^{(t-1)}$$

$$= \sum_{k=1}^{m} \lim_{n\to\infty} \frac{n\ (0)}{n} \cdot \sum_{t=1}^{T} p_{ki}^{(t-1)}$$

$$= \sum_{k=1}^{m} n_k \sum_{t=1}^{T} p_{ki}^{(t-1)}.$$

Then by a convergence theorem (see Cramer, 1946, p.254)

$\sqrt{n}\,(\hat{p}_{ij} - p_{ij})$ has the same limit distribution as

$$\frac{\sum_{t=1}^{T} \left[n_{ij}\ (t) - p_{ij}\ n_i\ (t-1)\right] / n^{1/2}}{\sum_{k=1}^{m} \sum_{t=1}^{T} n_k\ p_{kj}^{\,[t-1]}} \quad . \tag{5.1}$$

Assume n_k (0) fixed. Then by arguments of the previous section we have

$$E\left[\sum_{t=1}^{T} \left[n_{ij}\ (t) - p_{ij}\ n_i\ (t-1)\right] / n^{1/2} \right] = 0,$$

$$E\left[\sum_{t}\ n_{ij}\ (t) - p_{ij}\ n_i\ (t-1)\right]^2 / n^{1/2}$$

$$= \sum_{k} \sum_{t} n_k\ (0)\ p_{ki}^{\,[t-1]}\, p_{ij}\ (1-p_{ij})/n,$$

$$E\left[\sum_{t} \left[n_{ij}\ (t) - p_{ij}\ n_i\ (t-1)\right] \left[\sum_{t} (n_{gh}\ (t) - p_{gh}\ n_g\ (t-1)\right] / n \right.$$

$$= -\delta_{ig} \sum_{k} \sum_{t} n_k\ (0)\ p_{ki}^{\,[t-1]}\, p_{ij}\ p_{gh}/n,$$

where δ_{ig} is the kronecker delta.

Let $\sum_{k} \sum_{t} n_k\ p_{ki}^{\,[t-1]} = \phi_i$.

Since $\sqrt{n}\ (\hat{p}_{ij} - p_{ij})$ has the same distribution as (5.1) the variables

$$\sqrt{n}\ (\hat{p}_{ij} - p_{ij}) \longrightarrow \text{ mean } 0, \text{ variance } p_{ij}\ (1-p_{ij})/\phi_i$$

$$\text{covariance } -\delta_{ig}\ p_{ij}\ p_{gh}/\phi_i.$$

$$\sqrt{n\phi_i}\ (\hat{p}_{ij} - p_{ij}) \rightarrow \text{ mean } 0, \text{ variance } p_{ij}\ (1-p_{ij})$$

$$\text{covariance } -\delta_{ig}\ p_{ij}\ p_{gh}$$

$$\sqrt{n_i^*}\ (\hat{p}_{ij} - p_{ij}) \rightarrow \text{ mean } 0, \text{ variance } p_{ij}\ (1-p_{ij})$$

$$\text{covariance } -\delta_{ig}\ p_{ij}\ p_{gh}$$

where $n_i^* = \sum_{t=0}^{T-1} n_i\ (t).$

(note; "$\rightarrow$" means "limiting joint normal distributed as").

The variables $(n\phi_i)^{1/2}\ (\hat{p}_{ij} - p_{ij})$ for $\underline{m}$ different values of $\underline{i}$ $(i=1,2,\cdots, m)$ are asymptotically independent, and hence have the same limiting joint distribution as obtained from similar functions of the estimates of multinomial probabilities p_{ij} from $\underline{m}$ independent samples with sample size $n\phi_i$ $(i=1,2,\cdots, m)$.

However, the variable $\hat{p}_{ij}\ (t) = n_{ij}\ (t)/n_i\ (t-1)$ for a given $\underline{i}$ and $\underline{t}$ have the same asymptotic distribution as the estimates of multinomial probabilities with sample sizes $E\ (n_i\ (t-1)$, and the variables $\hat{p}_{ij}\ (t)$ for two different values of $\underline{i}$ or $\underline{t}$ are asymptotically independent.

TESTS OF HYPOTHESES

Test of Hypotheses about Specific Probabilities

Let p^o_{ij} $(i,j = 1,2,\cdots,m)$ be given values and consider the problem of testing $Ho : p_{ij} = p^o_{ij}$, $j = 1,2,\cdots,m$, for a given i against $Ha : p_{ij} = p^o_{ij}$, all i and j. Under Ho,

$$\sum_{j=1}^{m} n^*_i \frac{(\hat{p}_{ij} - p^o_{ij})^2}{p^o_{ij}} \tag{3.1.1}$$

is asymptotically χ^2_{m-1}. Since $n^*_i\ (\hat{p}_{ij} - p_{ij})^2$ for different i are asymptotically independent, the forms (3.1.1) for different i are asymptotically independent and hence can be added to obtain other χ^2-variables. For instance, a test for all p_{ij} $(i,j = 1,2,\cdots,m)$ can be obtained by adding (3.1.1) over all i, which results in a χ^2-variable with $m(m-1)$ d·f.

Testing Ho that the Transition Probabilities are Constant

To test $Ho : p_{ij}(t) = p_{ij}$ $(t = 1,2,\cdots,T)$, the estimates of the transition probabilities for time t are

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t-1)} .$$

Then the likelihood function maximized under Ho is

$$L_{max}\omega = \prod_t \prod_{i,j} \hat{p}_{ij}^{\,n_{ij}(t)} .$$

The likelihood function maximized under Ha is

$$L_{max}\Omega = \prod_{t} \prod_{i,j} p_{ij}(t)^{n_{ij}(t)} \quad .$$

The familiar likelihood ratio criterion is

$$\lambda = \prod_{t} \prod_{i,j} \left[\frac{\hat{p}_{ij}}{\hat{p}_{ij}(t)}\right]^{n_{ij}(t)}$$

and $-2 \log \lambda$ is distributed as $\chi^2_{(T-1)m(m-1)}$ when Ho is true. (Neyman, 1949)

An mxT contingency table can be used to represent the joint estimates $\hat{p}_{ij}(t)$ for a given $\underline{i}$ and for $\underline{j} = 1, 2, \cdots, m$, and $t = 1, 2, \cdots, T$. Thus

dependent →

independent ↑	1	2	···	m
1	$\hat{p}_{i1}(1)$	$\hat{p}_{i2}(1)$	···	$\hat{p}_{im}(1)$
2	$\hat{p}_{i1}(2)$	$\hat{p}_{i2}(2)$	···	$\hat{p}_{im}(2)$
·				
·				
·				
T				

and for each row there are $\underline{m}$ constants $p_{i1}, p_{i2}, \cdots, p_{im}$ with $\sum_{j} \hat{p}_{ij} = 1$.

The χ^2-test of homogeneity is given by

$$\chi_i^2 = \sum_{t,j} n_i(t-1)\,[\hat{p}_{ij}(t) - \hat{p}_{ij}]^2/p_{ij}$$

and χ_i^2 has $(m-1)(T-1)$ d.f., where $i = 1, 2, \cdots, m$.

Another test of the hypothesis that the transition probabilities are constant for T independent samples from multinomial trials can be obtained by using the likelihood ratio criterion. Thus

$$\lambda_i = \prod_{t,j} [\hat{p}_{ij}/\hat{p}_{ij}(t)]^{n_{ij}(t)} \,. \qquad (3.2.2)$$

The asymptotic distribution of $-2\log\lambda_i$ is χ^2 with $(m-1)(T-1)$ d.f., since it is related to the contingency table approach dealt with for a given <u>i</u>. Hence, Ho can be tested separately for each value of <u>i</u>.

Consider the joint hypothesis that $p_{ij}(t) = p_{ij}$ for all $i, j = 1, 2, \cdots, m$, $t = 1, 2, \cdots, T$. A test of this joint Ho can obtained from $\hat{p}_{ij}(t)$ and p_{ij} directly since the χ_i^2's are asymptotically independent. Hence

$$\chi^2_{m(m-1)(T-1)} = \sum_{i=1}^{m} \chi_i^2 = \sum_{i} \sum_{t,j} n_i\,(t-1)[\hat{p}_{ij}(t) - \hat{p}_{ij}]^2\,/\hat{p}_{ij} \qquad (3.3.3)$$

and the test criterion based on (3.2.2) is

$$\sum_{i=1}^{m} -2\log\lambda_i = -2\log\lambda.$$

Test of the Hypothesis that the Chain is of a Given Order

A 1st-order stationary chain is a special 2nd-order chain, for which $p_{ijk}(t)$ does not depend on i. Thus a 2nd-order chain can be represented as a more complicated 1st-order chain. To do this, let the pair of successive states i and j define a composite state (i,j). Then, the probability of the composite state (j,k) at t given the composite state (i,j) at t-1 is $p_{ijk}(t)$. The probability of state (h,k), h = j, given (i,j), is zero.

Assume $n_i(0)$ and $n_{ij}(1)$ are nonrandom. Consider the set $n_{ijk}(t)$ (i, j, k = 1, 2, $\cdots$, m; t = 2, 3, $\cdots$, T). The conditional distribution of $n_{ijk}(t)$, given $n_{ij}(t-1)$, is

$$\frac{n_{ij}(t-1)!}{\Pi_k\, n_{ijk}(t)!} \prod_{k=1}^{m} p_{ijk}^{\,n_{ijk}(t)}$$

where $n_{ij}(t-1) = \sum_k n_{ijk}(t)$.

The joint distribution of $n_{ijk}(t)$ for i, j, k = 1, 2, $\cdots$, m and t = 2, 3, $\cdots$, T is

$$\prod_{t=2}^{T} \prod_{i,j}^{m} \frac{n_{ij}(t-1)!}{\prod_k n_{ijk}(t)!} \prod_{k=1} p_{ijk}^{n_{ijk}(t)} .$$

The maximum likelihood estimate of p_{ijk} for stationary chains is

$$\hat{p}_{ijk} = \frac{n_{ijk}}{\sum_{t=1}^{m} n_{ijl}} = \frac{\sum_{t=2}^{T} n_{ijk}(t)}{\sum_{t=2}^{T} n_{ij}(t-1)} .$$

Consider Ho : $p_{1jk} = p_{2jk} = \cdots = p_{jk}$, for all $j, k = 1, 2, \cdots, m$. The likelihood criterion for testing this hypothesis is

$$\lambda = \prod_{i,j,k}^{m} (\hat{p}_{jk} / p_{ijk})^{n_{ijk}} \tag{3.3.1}$$

where $\hat{p}_{jk} = \sum_i n_{ijk} / \sum_i \sum_l n_{ijl}$

$$= \sum_{t=2}^{T} n_{jk}(t) / \sum_{t=1}^{T-1} n_j(t) .$$

Under Ho, $-2 \log\lambda$ is asymptotically χ^2 $m(m-1)^2$.

In contingency tables, for a given j, the $n^{1/2}$ $(\hat{p}_{ijk} - p_{ijk})$ have the same asymptotic distribution as the estimates of multinomial probabilities for m independent samples (i = 1, 2, ···, m). An mxm table can be used to represent $\hat{p}_{ijk}$ for a given j and for i, k = 1, 2, ···, m. To test

$$Ho : p_{ijk} = p_{jk} \qquad \text{for } i = 1, 2, \cdots, m, \text{ we have}$$

$$\chi_j^2 = \sum_{i,k} n_{ij}^* (\hat{p}_{ijk} - \hat{p}_{jk})^2 / \hat{p}_{jk}$$

where
$$n_{ij}^* = \sum_k n_{ijk} = \sum_k \sum_{t=2}^{T} n_{ijk}(t)$$

$$= \sum_{t=2}^{T} n_{ij}(t-1) = \sum_{t=1}^{T-1} n_{ij}(t)$$

If Ho is true, χ_j^2 has the usual limiting distribution with $(m-1)^2$ d.f.

For the use of the likelihood ratio criterion to test Ho, we calculate

$$\lambda_j = \prod_{i,k} (\hat{p}_{jk} / \hat{p}_{ijk})^{n_{ijk}}$$

and $-2 \log \lambda_j$ is $\chi^2_{(m-1)^2}$.

Consider the joint hypothesis that $p_{ijk} = p_{jk}$ for all i, j, k = 1, 2,$\cdots$, m. To test this joint hypothesis, compute the sum

$$\chi^2 m(m-1)^2 = \sum_{j=1}^{m} \chi_j^2 = \sum_{i,j,k} n_{ij}^{*} (\hat{p}_{ijk} - \hat{p}_{jk})^2 / \hat{p}_{jk} .$$

The test criterion based on (3.3.1) can be written

$$\sum_{j}^{m} - 2 \log \lambda_j = -2 \log \lambda$$

$$= 2 \sum_{i,j,k} n_{ijk} \log [\hat{p}_{ijk} / \hat{p}_{jk}]$$

$$= 2 \sum_{i,j,k} n_{ijk} [\log \hat{p}_{ijk} - \log \hat{p}_{jk}].$$

Consider Ho : $p_{ij\cdots kl} = p_{j\ldots kl}$ for i = 1, 2, $\cdots$, m; that is, test the hypothesis that a chain is of order r-1 against the alternative that it is of order r. For this Ho let $n_{ij\cdots kl}(t)$ be the states i, j, $\cdots$, k, l at times t-r, t-r+1, $\cdots$, t-1, t respectively, and $n_{ij\cdots kl}$ (t-1) = $\Sigma_l n_{i\cdots kl}(t)$. Assume here that the $n_{ij\ldots k}$ (r-1) are nonrandom. The maximum likelihood estimate of $p_{ij\cdots kl}$ is

$$\hat{p}_{ij\cdots kl} = n_{ij\cdots kl} / n^{*}_{ij\cdots k}$$

where $$n_{ij\cdots kl} = \sum_{t=r}^{T} n_{ij\cdots kl}(t) \text{ and}$$

$$n^*_{ij\cdots k} = \sum_{l} n_{ij\cdots kl} = \sum_{t=r}^{T} n_{ij\cdots k}(t-1)$$

$$= \sum_{t=r-1} n_{ij\cdots k}(t).$$

For a given set j, ..., k, the set $\hat{p}_{ij\cdots kl}$ will have the same asymptotic distribution as estimates of multinomial probabilities for m independent samples (i = 1, 2, ..., m), and can represented by an mxm table. Thus to test Ho : $p_{ij\cdots kl} = p_{j\cdots kl}$ (for i = 1, 2, ..., m) is true, we have

$$\chi^2_{j\cdots k} = \sum_{i,l} n^*_{ij\cdots k} (\hat{p}_{ij\cdots kl} - \hat{p}_{j\cdots kl})^2 / \hat{p}_{j\cdots kl}$$

where $$\hat{p}_{j\cdots kl} = \sum_{i} n_{ij\cdots kl} / \sum_{i} n^*_{ij\cdots k}$$

$$= \sum_{t=r}^{T} n_{j\cdots kl}(t) / \sum_{t=r-1}^{T-1} n_{j\cdots k}(t).$$

The $\chi^2_{j\cdots k}$ has $(m-1)^2$ d.f.. Since there are m^{r-1} sets j, ..., k (j = 1, ..., m; ...; k = 1, 2, ..., m) then

$$\chi^2_{total} = \Sigma_{j,\cdots,k}\, x^2_{j\cdots k}$$

will have $m^{r-1}(m-1)^2$ d.f. under the joint null hypothesis. One could use the likelihood ratio criterion

$$\lambda_{j\cdots k} = \prod_{i,\cdots,l} (\hat{p}_{j\cdots k} / \hat{p}_{ij\cdots l})^{n_{ij\cdots kl}}$$

where $-2\log\lambda_{j\cdots k}$ is distributed asymptotic as χ^2 with $(m-1)^2$ d.f. as a basis for testing Ho. Also

$$\sum_{j\cdots k}\left\{-2\log\lambda_{j\cdots k}\right\} = 2 \sum_{i\cdots kl} n_{ij\cdots kl} \log (\hat{p}_{ij\cdots l} / \hat{p}_{j\cdots kl})$$

has a limiting χ^2-distribution with $m^{r-1}(m-1)^2$ d.f. when the joint Ho is true.

χ^2-tests and Likelihood Criterion

The following development for testing certain hypothesis about single chains is due to Bartlett (1951). Consider the observed sequence $x_1, x_2, \cdots, x_{n-1}, x_n$. The probability of this sequence s is

$$p\left\{S\right\} = p\left\{x_1\right\} p\left\{x_2|x_1\right\} p\left\{x_3|x_1, x_2\right\} \cdots p\left\{x_k|x_1, x_2, \cdots, x_{k-1}\right\}$$

$$\cdot \prod_{i=1}^{n-k} p\left\{x_{k+i}|x_i, \cdots, x_{k+i-1}\right\} \qquad (4.1)$$

The variable x can take s values as the states 1, 2, $\cdots$, s and hence a subsequence $x_h, x_{h+1}, \cdots, x_{k+h-1}, x_{k+h}$ can take s^{k+1} values. Let the frequency of length k+1 be $n_{ij\cdots qr}$. Let $n_{ij\cdots qr} = n_{ur}$ and $p_{ur} = P\left\{x_r|x_i, \cdots, x_q\right\}$ Then (4.1) may be written

$$L = \log p\left\{S\right\} = \sum_{j=1}^{k} \log p\left\{x_j|x_1, x_2, \cdots, x_{j-1}\right\} + \sum_{u,r} n_{ur} \log p_{ur}. \qquad (4.2)$$

If n increases, then $\Sigma_{u,r} n_{ur} \log p_{ur}$ will become the dominant part of $\log p[S]$. If n is large enough then under the condition $\sum_r p_{ur} = 1$, it follows that

$$\hat{p}_{ur} = \frac{n_{ij\cdots qr}}{n_{ij\cdots q}} = \frac{n_{ur}}{n_u} \qquad (4.3)$$

where $n_u = \Sigma_r n_{ur}$. Hence the likelihood criterion becomes

$$\lambda = -2\left[L - L_{max}\right] = -2\left[\sum_{u,r} n_{ur} \log p_{ur} - \sum_{u,r} n_{ur} \log \hat{p}_{ur}\right]$$

$$= -2 \sum_{u,r} n_{ur} \log (p_{ur}\, n_u / n_{ur})$$

$$= -2 \sum_{u,r} n_{ur} \log \left(\frac{m_{ur}}{np_u} \cdot \frac{n_u}{n_{ur}} \right)$$

$$= -2 \sum_{u,r} n_{ur} \log \left[\frac{m_{ur}}{n_{ur}} \cdot \frac{n_u}{m_u} \right]$$

$$= 2 \left\{ \sum_{u,r} n_{ur} \log \left(\frac{n_{ur}}{m_{ur}} \right) - \sum_u n_u \log \frac{n_u}{m_u} \right. , \qquad (4.4)$$

where $m_{ur} = nP_{ur} = np_u p_{ur}$, $m_u = np_u$, P_{ur} P_u denote absolute probabilities of the 'values' (u,r) and u.

In the case of k=0, we have independence and (4.2) becomes

$$L = \Sigma_r\, n_r \log p_r$$

and (4.4) becomes

$$\lambda = 2\, \Sigma_r\, n_r \log (n_r / m_r) \qquad (4.5)$$

The expression (4.5) is the likelihood criterion and is asymptotically a χ^2

distribution with (S-1) d.f. Alternatively, one could use

$$\chi^2_{-1} = \sum_r \frac{(n_r - m_r)^2}{m_r}$$

to test Ho : $p_{ij\cdots kl} = p_{j\cdots kl}$ $(i = 1, 2, \cdots, s)$.

Hoel (1954) has given another approach.

Let

$$L = \prod_{i,\cdots l} p_{ij\cdots kl}^{\,n_{ij\cdots kl}}$$

where $i, j, \cdots, k, l = 1, 2, \cdots, m$. Corresponding to m possible states, let $n_{ij\cdots kl}$ denote the frequency of the r-order chain state $ij\cdots kl$ for r+1 subscripts. Then, the maximum-likelihood estimate of $p_{ij\cdots kl}$ is

$$\hat{p}_{ij\cdots kl} = \frac{n_{ij\cdots kl}}{n_{ij\cdots k}}$$

where $n_{ij\cdots k} = \sum_l n_{ij\cdots kl}$.

Let $\hat{p}'_{ij\cdots kl} = \hat{p}_{j\cdots kl} = \dfrac{n_{j\cdots kl}}{n_{j\cdots k}}$ under Ho.

The likelihood ratio for testing Ho is

$$\lambda = \frac{L_o\ (\max)}{L\ (\max)} = \frac{L\ (\hat{p}'_{ij\cdots kl})}{L\ (\hat{p}_{ij\cdots kl})} .$$

Assume the $n_{ij\cdots kl}$ are asymptotic normally distribution.

Hence $$L \sim |A|^{\frac{1}{2}} e^{-\frac{1}{2}\ (n-u)A(n-u)'}$$

where (n-u) denotes the row vector of the linearly independent variables $n_{ij\cdots kl} - u_{ij\cdots kl}$, where $E(n_{ij\cdots kl}) = u_{ij\cdots kl}$ and A is positive definite matrix. Then

$$\lambda \sim \frac{|\hat{A}'|^{\frac{1}{2}} e^{-\frac{1}{2}\ (n-\hat{u}')\ \hat{A}'(\ n-\hat{u}')'}}{|\hat{A}\ |^{\frac{1}{2}} e^{-\frac{1}{2}\ (n-\hat{u})\ \hat{A}\ (n-\hat{u})'}}$$

where $\hat{u}$, $\hat{A}$ and $\hat{u}'$, $\hat{A}'$ indicate that the parameters $p_{ij\cdots kl}$ have been replaced by $\hat{p}_{ij\cdots kl}$ and $\hat{p}'_{ij\cdots kl}$, respectively. Now

$$-2 \log \lambda \text{ is approximately } \log \frac{|\hat{A}\ |}{|\hat{A}'|} + [n-\hat{u}]\ \hat{A}'\ [n-u']' - [n-\hat{u}\ ]\ \hat{A}\ [n-\hat{u}]'$$

when Ho is true. Then $|\hat{A}\,|$ and $|\hat{A}'|$ converges stochastically to the same value.

Hence

$$-2 \log \lambda \sim [n-\hat{u}'] \hat{A}' [n-u']' - [n-\hat{u}] \hat{A} [n-u].$$

Since $$u_{ij\cdots kl} = E(n_{ij\cdots kl}) = nP_{ij\cdots k}\, p_{ij\cdots kl}$$

where $P_{ij\cdots k}$ denotes the absolute probability of obtaining the r-1 chain state $ij\cdots k$, then

$$\hat{u}_{ij\cdots kl} = n\hat{P}_{ij\cdots k}\, \hat{p}_{ij\cdots kl}.$$

Since $P_{ij\cdots k}$ is some function of the transition probabilities $p_{ij\cdots kl}$, it can be written as

$$P_{ij\cdots k} = g\,(p_{ij\cdots kl})$$

and $$\hat{P}_{ij\cdots k} = g\,(\hat{p}_{ij\cdots kl}).$$

Assume that $g\,(\hat{p}_{ij\cdots kl}) = \dfrac{n_{ij\cdots k}}{n}$

then $$\hat{\mu}_{ij\cdots kl} = n\cdot \frac{n_{ij\cdots k}}{n} \cdot \frac{n_{ij\cdots kl}}{n_{ij\cdots k}} = n_{ij\cdots kl}$$

Therefore $\quad -2 \log \lambda \sim [n-\hat{u}'] \ \hat{A}' \ [n-u'].$ (4.6)

The right hand side of (4.6) is a quadratic form and is distributed as

$$x^2_{s^{r-1}} \ (s-1)^2$$

and hence $\quad -2 \log \lambda \sim x^2_{s^{r-1}} \ (s-1)^2.$

Anderson and Goodman (1957) have the following approach to χ^2-tests and the likelihood ratio criterion. Consider the distribution of the χ^2-statistics (3.3.3) under Ho : $p_{ij} \ (t) = p_{ij}$ for all $i, j = 1, 2, \cdots, m.$ $t = 1, 2, \cdots, T.$ Since $\sqrt{n} \ (\hat{p}_{ij}(t) - p_{ij})$ are asymptotically normally distributed with mean zero and variance $p_{ij} \ (1-p_{ij}) \ / \ m_i(t-1)$, etc., where

$$E \left[\frac{n_i(t)}{n} \right] = m_i(t)$$

then for different $\underline{t}$ or different $\underline{i}$, they are asymptotically independent. Then $[nm_i \ (t-1)]^{1/2} \ [p_{ij}(t) - p_{ij}] \sim N[o, \ p_{ij} \ (1-p_{ij})]$, etc.,

Let $\quad \hat{p}^*_{ij} = \Sigma_t \ m_i(t-1) \ \hat{p}_{ij}(t) \ / \ \Sigma_t \ m_i(t-1).$ Then

$$\Sigma_t \ nm_i(t-1) \left[\hat{p}_{ij}(t) - \hat{p}^*_{ij} \right]^2 \sim \chi^2 \quad \text{under Ho.}$$

But $$p \lim \hat{p}^*_{ij} = p \lim \Sigma_t\, m_i(t-1)\, \hat{p}_{ij}(t) \,/\, \Sigma_t\, m_i(t-1)$$

$$= p \lim \Sigma_t\, n_i(t-1)\, \hat{p}_{ij}(t) \,/\, \Sigma_t\, n_i(t-1)$$

$$= p \lim \Sigma_t\, n_i(t) \,/\, \Sigma_t\, n_i(t-1)$$

$$= \hat{p}_{ij}$$

and $$p \lim \left(\frac{n_i(t)}{n} - m\,(t)\right) = 0.$$

Therefore $$p \lim \left[n\, \Sigma \frac{m_i(t-1)\, (\hat{p}_{ij}(t) - p^*_{ij})^2}{\hat{p}^*_{ij}} \right]$$

$$= \Sigma_t \frac{n_i(t-1)\, (\hat{p}_{ij}(t) - \hat{p}_{ij})^2}{\hat{p}_{ij}} \quad .$$

Hence, the χ^2-statistics has the same asymptotic distribution as $\Sigma\, nm_i(t-1)\,[\hat{p}_{ij}(t) - \hat{p}^*_{ij}]^*$; that is, a χ^2-distribution.

Next, consider that for $|x|$ 1/2,

$$(1+x) \log (1+x) = (1+x) \left(x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \right)$$

$$= x + \frac{x^2}{2} - (x^3/6) \left(1 - \frac{x}{2} + \cdots \right)$$

and $\left| (1+x) \log (1+x) - x - \frac{x^2}{2} \right| = \left| (x^3/6)\,(1 - \frac{x}{2} + \cdots)\right| \le |x^3|$

since $\lambda_i = \prod_{t,j} \left[\hat{p}_{ij} / p_{ij}(t) \right]^{n_{ij}(t)}$.

Therefore $-2 \log \lambda_i = -2 \sum_{t,j} n_{ij}(t) \log \hat{p}_{ij} / \hat{p}_{ij}(t)]$

$$= 2 \sum_{j,t} n_i(t-1)\, \hat{p}_{ij} \log[\hat{p}_{ij}(t) / \hat{p}_{ij}]$$

$$= 2 \sum_{t,j} n_i(t-1)\, \hat{p}_{ij}[1 + x_{ij}(t)] \log[1 + x_{ij}(t)]$$

where $x_{ij}(t) = [\hat{p}_{ij}(t) - \hat{p}_{ij}] / p_{ij}$.

Then $\Delta = -2 \log \lambda_i - x_i^2$

$$= 2 \sum_{j,t} n_i(t-1)\, \hat{p}_{ij} \left\{ 1 + x_{ij}(t) | \log [1 + x_{ij}(t)] - \frac{[x_{ij}(t)]^2}{2} \right\}$$

$$= 2 \sum_{j,t} n_i(t-1)\, \hat{p}_{ij} \left\{ [1 + x_{ij}(t)] \log [1 + x_{ij}(t)] - x_{ij}(t) - \frac{(x_{ij}(t))^2}{2} \right\}$$

since $\Sigma_j\, \hat{p}_{ij}\, x_{ij}(t) = \Sigma_j\, (\hat{p}_{ij}(t) - \hat{p}_{ij}) = 0$

for any $\varepsilon > 0$, we have

$$p_r\left\{|\Delta| < \varepsilon\right\} \geq p_r\left\{|\Delta| < \varepsilon \text{ and } |x_{ij}(t)| < \frac{1}{2}\right\}$$

$$\geq p_r\left\{2 \sum_{j,t} n_i(t-1)\, \hat{p}_{ij}\left[x_{ij}(t)\right]^3 < \varepsilon \text{ and } \left|x_{ij}(t)\right| < \frac{1}{2}\right\}$$

$$\geq p_r\left\{2n \sum_{j,t} \left|x_{ij}(t)\right|^3 \leq \varepsilon \text{ and } |x_{ij}(t)| < \frac{1}{2}\right\}$$

since $\quad p \lim x_{ij}(t) = p \lim \;\; \hat{p}_{ij}(t) - \hat{p}_{ij} \;\; / \; \hat{p}_{ij} = 0.$

Therefore $\quad p \lim n\left[x_{ij}(t)\right]^3 = p \lim \left[(x_{ij}(t)\; n)^{1/2} \cdot x_{ij}(t)\right]^2$

$$= p \lim \sqrt{x_{ij}(t)\; n}\; \left\{\frac{\hat{p}_{ij}(t) - \hat{p}_{ij}}{\hat{p}_{ij}} - \frac{\hat{p}_{ij} - p_{ij}}{\hat{p}_{ij}}\right\}$$

$$= 0$$

Hence $\quad p_r\left\{|\Delta| < \varepsilon\right\} = 0 \text{ and } -2\log \lambda_i \sim \chi_i^2 .$

Application and Example

To illustrate the usefulness of the theoretical results discussed in the previous sections, we consider an example from climatology (Feyerherm and Bark, 1964). Consider the problem of testing hypothesis concerning the order of a Markov chain composed of a sequence of wet and dry days. We assume that

$p_{ij}(t) = p_{ij}$, $t = 1, 2, \cdots, T$, if T is less than 41 days and that successive years can be considered as repeated observations on the same chain.

The test statistic is easy to compute from ordinary contingency tables which show observed numbers for various cells of the table. Data for Manhattan, Kansas for the 40-days period begining on the 7th day and ending with the 46th day of the year were as shown in Table 1-4, where the states are taken to be D (dry day) and W (wet day).

Table 1. Observed values for testing

$Ho : p_{jk} = p_k$ vs $Ha : p_{jk} \neq p_k$, $j,k = D,W$.

	k=D	k=W	
j=D	1799	262	2061
j=W	261	118	379
	2060	380	2440

$\chi_1^2 = 82.631$

Table 2. Observed values for testing

$Ho : p_{ijk} = p_{jk}$ vs $Ha : p_{ijk} \neq p_{jk}$, $i,j,k = D,W.$

t-1 \ t, t-2	j=D k=D	j=D k=W	
i=D	1579	229	1808
i=W	220	33	253
	1799	262	2061

t-1 \ t, t-2	j=W k=D	j=W k=W	
i=D	178	83	261
i=W	83	35	118
	261	118	379

$\chi^2_{D,1} = .0285$

$\chi^2_{W.1} = .1735$

$$\chi^2_2 = x^2_D + x^2_W = .2020$$

From Table 1, the hypothesis of independence (chain is of zero order) is firmly rejected. For the hypothesis that the chain is of order two rather than one (table 2) three χ^2 values were computed. The first is one for sequences in which the middle day ($(t-1)^{st}$ day) was dry, the second for sequences in which the middle day was wet and the third is the sum of the first two. They provide asymptotically independent tests of the same Ho.

which could be stated alternatively as: Given the weather (D or W) on the middle day of a three-day sequences; the weather (D or W) on the 3rd day is independent of the weather (D or W) on the 1st day. Evidence for rejecting Ho is insufficient.

Before stating that we are dealing with a first-order chain we might look at some other hypothesis.

Table 3. Observed values for testing

$Ho : p_{ijkl} = p_{jkl}$ vs $Ha : p_{ijkl} \neq p_{jkl}$, $i,j,k,l = D,W$

t-3 \ t-2, t-1 / t	j,k = D,D l=D	j,k = D,D l=W	
i=D	1389	190	1597
i=W	200	29	229
	1589	219	1808

t-3 \ t-2, t-1 / t	j,k = D,W l=D	j,k = D,W l=W	
i=D	158	20	178
i=W	71	12	83
	229	32	261

$\chi^2_{DD,1} = .075$

$\chi^2_{DW,1} = .546$

t-2, t-1 / t / t-3	j,k = W,D 1=D	1=W	
i=D	155	65	220
i=W	18	15	33
	173	80	253

t-2, t-1 / t / t-3	j,k = W,W 1=D	1=W	
i=D	59	24	83
i=W	26	9	35
	85	33	118

$\chi^2_{WD,1} = 3.359$

$\chi^2_{WW.1} = .125$

$$\chi^2_4 = \chi^2_{DD,1} + \chi^2_{DW,1} + \chi^2_{WD,1} + \chi^2_{WW,1} = 4.105$$

Alternatively the hypothesis in Table 3 can be stated: Given the weather (DD, DW, WD, or WW) on the middle 2-days of a 4-day sequence, the weather (D or W) on the 4th day is independent of the weather (D or W) on the 1st day. Again, there is no basis for rejecting Ho.

Table 4. Observed values for testing

Ho : $p_{ijkl} = p_{kl}$ vs Ha : $p_{ijkl} \neq p_{kl}$, i,j,k,l = D,W

t-1 \ t / t-3, t-2	k=D l=D	k=D l=W	
i,j=DD	1389	200	1589
i,j=DW	155	18	173
i,j=WD	190	29	219
i,j=WW	65	15	80
	1799	262	2061

t-1 \ t / t-3, t-2	k=W l=D	k=W l=W	
i,j=DD	158	71	229
i,j=DW	59	26	85
i,j=WD	20	12	32
i,j=WW	24	9	33
	261	118	379

$\chi^2_{D,3} = 3.536$

$\chi^2_{W,3} = .848$

$$\chi^2_6 = \chi^2_{D,3} + \chi^2_{W,3} = 4.384$$

Alternatively, the hypothesis in Table 4 can be stated: Given the weather (D or W) on the 3rd day of a 4-day sequence, the weather (D or W) on the 4th-day is independent of the weather (D or W) on the 1st two days. Again, there is no basis for rejecting Ho. The results in Tables (1-4) indicate that a first order chain represents a good approximation for describing the dependence in a sequence of wet and dry days.

ACKNOWLEDGMENT

The writer wishes to express his sincere appreciation to his major professor, Dr. Arlin M. Feyerherm, who gave helpful suggestions and assistance during the preparation of this report and to Dr. H. C. Fryer and all the faculty members in the Department of Statistics for their continuous encouragement and instruction through his graduate study.

REFERENCES

Anderson, T. W., "Probability models for analyzing times changes in attitudes", Mathematical Thinking in the Social Sciences, Illinois. (1954)

Anderson, T. W. and Goodman, L. A., "Statistical inference about Markov chains", Ann. Math. Stat., 28: 89-109. (1957)

Bartlett, M. S., "The frequency goodness of fit test for probability chains", Proc. Cambridge Soc., 47: 86-95. (1951)

Cramer, H., Mathematical Methods of Statistics, Princeton Univ. Press, Princeton. (1946)

Feller, W., An Introduction to Probability Theory and its Applications, John Wiley and Sons Inc. (1957)

Feyerherm, A. M. and Bark, L. D., "Statistical Methods for Persistant Precipitation Patterns", Paper presented at Sixth National conference on Agricultural Meterology, October 8-10. (1964)

Gabriel, K. R., and Neumann, J., On a distribution of weather cycles by length. Quart. J. R. Meteor. Soc., 83: 375-380. (1957)

Good, I. J., "The likelihood ratio test for Markov chain", Biometrika, 42: 531-3. (1955) ; Corrigenda, 44:301. (1957)

Hoel, P. G., "A test for Markov chains", Biometrika, 41: 430-433. (1954)

Hogg, R. V. and Craig, A. T., Introduction to Mathematical Statistics, 2nd Ed., The Macmillon Company, New York. (1965)

Neyman, J., "Contribution to the theory of the x^2-test", Proceedings of the Berkeley Symposium on Math. Stat. and Probability, Univ. of Cali. Press, Berkeley, 239-274. (1949)

Parzen, E., Stochastic Process, Holden-Day Inc. (1962)

Wilks, S. S., Mathematical Statistics, John Wiley and Sons, Inc. (1963)

STATISTICAL INFERENCE ABOUT MARKOV CHAINS

by

SAI-SING LIN

B. Sc., Taiwan Normal University, 1958

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirement for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1966

The determination of limiting distribution functions of certain functions of n random variables as $n \to \infty$ is an important class of problems in mathematical statistics.

In Markov chains, the maximum likelihood estimates and their asymptotic distribution are obtained for the transition probabilities in a chain of arbitrary order when there are repeated observation of the same chain.

Likelihood ratio tests and χ^2-tests of the form used in contingency tables are obtained for testing the following hypotheses: (a) P_t is stationary (i.e. ; $P_t = P = \left\{ p_{ij} \right\}$) against the alternative that it varies over time, (b) P is a given matrix against the alternative that it is not, (c) the process is a u^{th} order Markov chain against the alternative it is r^{th} but not u^{th} order. In case u=0 and r=1, case (c) results in tests of the null hypothesis that observations at successive time points are statistically independent against the alternate hypothesis that observation are from a first order Markov chain.

There is some disscusion of the relation between the likelihood ratio criterion and χ^2-tests of the form used in contingency tables. An example which shows the usefulness of the theory is given.